Statistics in Medicine

Calculating confidence intervals for some non-parametric analyses

MICHAEL J CAMPBELL,   MARTIN J GARDNER

Gardner and Altman[1] described the rationale behind the use of confidence intervals and gave methods for their calculation for a population mean and for differences between two population means for paired and unpaired samples. These methods are based on sample means, standard errors, and the t distribution and should strictly be used only for continuous data from Normal distributions (although small deviations from Normality are not important[2]).

For non-Normal continuous data the median of the population or the sample is preferable to the mean as a measure of location. Medians are also appropriate in other situations—for example, when measurements are on an ordinal scale.

This paper describes methods of calculating confidence intervals for a population median or for other population quantiles from a sample of observations. Calculations of confidence intervals for the difference between two population medians or means (a non-parametric approach rather than the parametric approach mentioned above) for both unpaired and paired samples are described. Worked examples are given for each situation.

Because of the discrete nature of some of the sampling distributions involved in non-parametric analyses it is not usually possible to calculate confidence intervals with exactly the desired level of confidence. Hence, if a 95% confidence interval is wanted the choice is between the lowest possible level of confidence over 95% (a "conservative" interval) and the highest possible under 95%. There is no firm policy on which of these is preferred, but we will mainly describe conservative intervals in this paper. The exact level of confidence associated with any particular approximate level can be calculated from the distribution of the statistic being used.

The methods outlined for obtaining confidence intervals are described in more detail in textbooks on non-parametric statistics.[3] The calculations can be carried out using the statistical computer package MINITAB.[4] A method for calculating confidence intervals for Spearman's rank correlation coefficient is given in an accompanying paper.[5]

A confidence interval indicates the precision of the sample statistic as an estimate of the overall population value. Confidence intervals convey the effects of sampling variation but cannot control for non-sampling errors in study design or conduct. They should not be used for basic description of the sample data but only for indicating the uncertainty in sample estimates for population values of medians or other statistics.

Confidence intervals for medians and other quantiles

MEDIAN

The median is defined as the value having half of the observations less than and half exceeding it. The sample median is used as an estimate of the population median. To find the $100(1-\alpha)\%$ confidence interval for the population median first calculate the quantities:

$$r = \frac{n}{2} - \left(N_{1-\alpha/2} \times \frac{\sqrt{n}}{2} \right) \quad \text{and} \quad s = 1 + \frac{n}{2} + \left(N_{1-\alpha/2} \times \frac{\sqrt{n}}{2} \right),$$

where n is the sample size and $N_{1-\alpha/2}$ is the appropriate value from the standard Normal distribution for the $100(1-\alpha/2)$ percentile. This is widely available in tables. Then round r and s to the nearest integers. The n sample observations need to be ranked in increasing order of magnitude and the rth to sth observations in the ranking then determine the $100(1-\alpha)\%$ confidence interval for the population median. This approximate method is satisfactory for most sample sizes.[6] The exact method based on the binomial distribution can be used instead as shown in the example below.

OTHER QUANTILES

A similar approach can be used to calculate confidence intervals for quantiles other than the median—for example, the upper 95th percentile. For the qth quantile (q=0·95 for the 95th percentile) r and s above are replaced with r' and s' given by:

$$r' = nq - (N_{1-\alpha/2} \times \sqrt{nq(1-q)})$$

and

$$s' = 1 + nq + (N_{1-\alpha/2} \times \sqrt{nq(1-q)}).$$

WORKED EXAMPLES

(1) In the example given by Gardner and Altman[1] the median systolic blood pressure among 100 diabetic patients was 146 mm Hg. Using the above formulae to calculate a 95% confidence interval gives:

$$r = \frac{100}{2} - \left(1·96 \times \frac{\sqrt{100}}{2} \right) = 40·2 \quad \text{and} \quad s = 1 + \frac{100}{2} + \left(1·96 \times \frac{\sqrt{100}}{2} \right) = 60·8.$$

From the original data the 40th observation in increasing order is 142 mm Hg and the 61st is 150 mm Hg. The 95% confidence interval for the population median is thus from 142 to 150 mm Hg.

(2) For a small sample example using the binomial distribution consider the results of a study measuring β endorphin concentrations in 11 subjects

Medical Statistics and Computing, University of Southampton, Southampton General Hospital, Southampton SO9 4XY

MICHAEL J CAMPBELL, PHD, senior lecturer in medical statistics

MRC Environmental Epidemiology Unit (University of Southampton), Southampton General Hospital, Southampton SO9 4XY

MARTIN J GARDNER, PHD, professor of medical statistics

Correspondence to: Dr Campbell.

who had collapsed while running in a half marathon.[7] The concentrations in pmol/l in order of increasing value were: 66·0, 71·2, 83·0, 83·6, 101, 107·6, 122, 143, 160, 177, and 414.

The sample median is the 6th observation (107·6 pmol/l). To find a confidence interval for the population median we use the binomial distribution[2] with probability of 0·5. For a conservative 95% confidence interval we first find the largest cumulative probability under 0·025 and the smallest over 0·975 either by direct calculation or from tables.[3] With n=11 this gives Prob (X≤1)=0·006 and Prob (X≤9)=0·994.

The approximate 95% confidence interval is then found by the ranked observations that are one greater than those associated with the two probabilities—that is, the 2nd and 10th observation, giving 71·2 to 177 pmol/l. The actual probability associated with this confidence interval is 0·994−0·006=0·988, so effectively it is a 98·8% confidence interval.

Alternatively a non-conservative approximate 95% confidence interval can be found by calculating the smallest cumulative probability over 0·025 and the largest under 0·975. In this case P(X≤2)=0·033 and P(X≤8)=0·967, which give a 93·4% confidence interval from the 3rd to the 9th ranked observations (83·0 to 160 pmol/l). In this case the coverage probability is nearer to 95% than for the conservative interval.

The exact level of confidence associated with any particular approximate level of confidence can be calculated from the distribution of the statistic being used.

Confidence intervals for differences

TWO SAMPLES

In finding confidence intervals for population differences it is assumed that the data come from distributions that are identical except in location. Because of this assumption the non-parametric confidence intervals described below can be regarded as being either for the difference between the two medians, or the difference between the two means, or the difference between any other two measures of location such as a particular percentile. This assumption is not necessary for a valid test of the null hypothesis of no difference in population distributions but if it is not satisfied the interpretation of a statistically significant result is difficult.

Unpaired case

Let $x_1, x_2, \ldots, x_n$ represent the n observations in a sample from one population and $y_1, y_2, \ldots, y_m$ the m observations in a sample from a second population, where both sets of data are thought not to come from Normal distributions. The difference between the two population medians or means is estimated by the median of all the possible $n \times m$ differences $x_i - y_j$ (for i=1 to n and j=1 to m).

The confidence interval for the difference between the two population medians or means is also derived through these $n \times m$ differences.[3] For an approximate $100(1-\alpha)\%$ confidence interval first calculate:

$$K = W_{\alpha/2} - \frac{n(n+1)}{2},$$

where $W_{\alpha/2}$ is the $100\alpha/2$ percentile of the distribution of the Mann-Whitney test statistic (see table A7 of Conover[3]) or of the distribution of the equivalent Wilcoxon two sample test statistic (see pp 156-62 of Lentner[8]). The Kth smallest to the Kth largest of the $n \times m$ differences then determine the $100(1-\alpha)\%$ confidence interval. Values of K for finding approximate 95% confidence intervals ($\alpha=0·05$) are given directly for sample sizes of up to 20 in Appendix 1.

For studies where each sample size is greater than about 20 K can be calculated approximately[3] as:

$$K = \frac{nm}{2} - \left(N_{1-\alpha/2} \times \sqrt{\frac{nm(n+m+1)}{12}} \right),$$

rounded up to the next integer value, where $N_{1-\alpha/2}$ is the appropriate value from the standard Normal distribution for the $100(1-\alpha/2)$ percentile.

Worked example—Consider the data on the globulin fraction of plasma (g/l) in two groups of 10 patients given by Swinscow[9]:

| Group 1 | 38 | 26 | 29 | 41 | 36 | 31 | 32 | 30 | 35 | 33 |
| Group 2 | 45 | 28 | 27 | 38 | 40 | 42 | 39 | 39 | 34 | 45 |

The computations are made easier if the data in each group are first ranked into increasing order of magnitude and then all the differences for group 1−group 2 calculated as in the following table:

| | Group 1 | | | | | | | | | |
Group 2	26	29	30	31	32	33	35	36	38	41
27	−1	2	3	4	5	6	8	9	11	14
28	−2	1	2	3	4	5	7	8	10	13
34	−8	−5	−4	−3	−2	−1	1	2	4	7
38	−12	−9	−8	−7	−6	−5	−3	−2	0	3
39	−13	−10	−9	−8	−7	−6	−4	−3	−1	2
39	−13	−10	−9	−8	−7	−6	−4	−3	−1	2
40	−14	−11	−10	−9	−8	−7	−5	−4	−2	1
42	−16	−13	−12	−11	−10	−9	−7	−6	−4	−1
45	−19	−16	−15	−14	−13	−12	−10	−9	−7	−4
45	−19	−16	−15	−14	−13	−12	−10	−9	−7	−4

The estimate of the difference in population medians or means is now given by the median of these differences. From the 100 differences in the table the 50th smallest difference is −6 g/l and the 51st is −5 g/l so the median difference is estimated as $(-6+(-5))/2 = -5·5$ g/l.

To calculate an approximate 95% confidence interval for the difference in population medians or means the value of K=24 is found for n=10, m=10, and $\alpha=0·05$ from the appropriate table or from Appendix 1. The 24th smallest difference is −10 g/l and the 24th largest is +1 g/l. The approximate 95% confidence interval (exact level 95·7%) for the difference in population medians or means is from −10 g/l to +1 g/l.

Paired case

Paired cases include studies of repeated measurements on the same individuals and matched case-control comparisons where paired differences are the observations of main interest. Suppose that in a sample of size n the differences for each matched pair of measurements are $d_1, d_2, \ldots, d_n$. The difference between the two population medians or means is estimated by calculating all the $n(n+1)/2$ possible averages of two of these differences taken together, including each difference with itself, and selecting their median.

The confidence interval for the difference between the population medians or means is also derived using these averaged differences. For an approximate $100(1-\alpha)\%$ confidence interval first find the value of $W_{\alpha/2}$ as the $100\alpha/2$ percentile of the distribution of the Wilcoxon one sample test statistic (see table A13 of Conover[3] or p 163 of Lentner[8]). Then if $W_{\alpha/2}=K^\star$ the $K^\star$th smallest to the $K^\star$th largest of the averaged differences determine the $100(1-\alpha)\%$ confidence interval. Values of $K^\star$ for finding approximate 95% confidence intervals ($\alpha=0·05$) are given for sample sizes of up to 20 in Appendix 2 together with the associated exact levels of confidence. In general the degree of conservatism over 95% confidence is small.

For sample sizes of about 20 or more special tables are not required and $K^\star$ can be calculated approximately[3] as:

$$K^\star = \frac{n(n+1)}{4} - \left(N_{1-\alpha/2} \times \sqrt{\frac{n(n+1)(2n+1)}{24}} \right),$$

rounded up to the next integer value, where $N_{1-\alpha/2}$ is the appropriate value from the standard Normal distribution for the $100(1-\alpha/2)$ percentile.

Worked example—Consider further results of measuring β endorphin concentrations in subjects running in a half marathon where 11 people were studied both before and after the event.[7] The before and after concentrations (pmol/l) and their differences ordered by increasing size were as follows:

| | β endorphin concentration | | Change in concentration |
Subject No	Before	After	After−before
1	10·6	14·6	4·0
2	5·2	15·6	10·4
3	8·4	20·2	11·8
4	9·0	20·9	11·9
5	6·6	24·0	17·4
6	4·6	25·0	20·4
7	14·1	35·2	21·1
8	5·2	30·2	25·0
9	4·4	30·0	25·6
10	17·4	46·2	28·8
11	7·2	37·0	29·8

All the possible n(n+1)/2 averages in this case where n=11 give the 66 averages shown below:

Change	4·0	10·4	11·8	11·9	17·4	20·4	21·1	25·0	25·6	28·8	29·8
4·0	4·00	7·20	7·90	7·95	10·70	12·20	12·55	14·50	14·80	16·40	16·90
10·4		10·40	11·10	11·15	13·90	15·40	15·75	17·70	18·00	19·60	20·10
11·8			11·80	11·85	14·60	16·10	16·45	18·40	18·70	20·30	20·80
11·9				11·90	14·65	16·15	16·50	18·45	18·75	20·35	20·85
17·4					17·40	18·90	19·25	21·20	21·50	23·10	23·60
20·4						20·40	20·75	22·70	23·00	24·60	25·10
21·1							21·10	23·05	23·35	24·95	25·45
25·0								25·00	25·30	26·90	27·40
25·6									25·60	27·20	27·70
28·8										28·80	29·30
29·8											29·80

Thus having found K*=11 for n=11 and α=0·05 from appropriate tables or from Appendix 2, the 11 smallest averages are 4·0, 7·2, 7·9, 7·95, 10·4, 10·7, 11·1, 11·15, 11·8, 11·85, and 11·9; and the 11 largest averages are 25·1, 25·3, 25·45, 25·6, 26·9, 27·2, 27·4, 27·7, 28·8, 29·3, and 29·8. The approximate 95% (exact 95·8%) confidence interval for the difference between the population medians or means is thus given as 11·9 to 25·1 pmol/l around the sample median which is 18·825 pmol/l (the average of the 33rd and 34th ranked observations, 18·75 and 18·90, in the table of average differences). The triangular table of average differences helps to identify the required values, but a computer package such as MINITAB[4] can rank the averages in order and select the appropriate ranked values.

It should be noted that there are differences of presentation in the tables referred to in Conover,[3] Geigy Scientific Tables,[8] and elsewhere. These result from the discrete nature of the distributions and whether[8] or not[3] the tabulated values are part of the critical region for the test of the null hypothesis. MINITAB[4] uses large sample formulae with continuity corrections for computing the coverage probabilities of the confidence intervals for differences even for sample sizes less than 20. This can lead to inaccuracies in the coverage probabilities given by the program.

differences in population medians of two paired samples with sample size n from 6 to 20 and the associated exact levels of confidence

n	K*	Level of confidence (%)
6	1	96·9
7	3	95·3
8	4	96·1
9	6	96·1
10	9	95·1
11	11	95·8
12	14	95·8
13	18	95·2
14	22	95·1
15	26	95·2
16	30	95·6
17	35	95·5
18	41	95·2
19	47	95·1
20	53	95·2

References

1 Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. Br Med J 1986;292:746-50.
2 Bland M. An introduction to medical statistics. Oxford: University Press, 1987:179-82.
3 Conover WJ. Practical non-parametric statistics. New York: Wiley, 1980.
4 Ryan BF, Joiner BL, Ryan TA. Minitab handbook. 2nd ed. Boston: Duxbury Press, 1985.
5 Altman DG, Gardner MJ. Calculating confidence intervals for regression and correlation. Br Med J 1988;296:1238-42.
6 Hill ID. 95% confidence limits for the median. J Statist Comput and Simulation 1987;28:80-1.
7 Dale G, Fleetwood JA, Weddell A, Ellis RD, Sainsbury JRC. β Endorphin: a factor in "fun run" collapse. Br Med J 1987;294:1004.
8 Lentner C, ed. Geigy scientific tables. Vol 2. 8th ed. Basle: Ciba-Geigy, 1982.
9 Swinscow TDV. Statistics at square one. 7th ed. London: British Medical Association, 1980.

Appendix 1:

Values of K for finding approximate 95% confidence intervals for differences in population medians of two unpaired samples with sample sizes n and m from 5 to 20.

n	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
5	3	4	6	7	8	9	10	12	13	14	15	16	18	19	20	21
6	4	6	7	9	11	12	14	15	17	18	20	22	23	25	26	28
7	6	7	9	11	13	15	17	19	21	23	25	27	29	31	33	35
8	7	9	11	14	16	18	20	23	25	27	30	32	35	37	39	42
9	8	11	13	16	18	21	24	27	29	32	35	38	40	43	46	49
10	9	12	15	18	21	24	27	30	34	37	40	43	46	49	53	56
11	10	14	17	20	24	27	31	34	38	41	45	48	52	56	59	63
12	12	15	19	23	27	30	34	38	42	46	50	54	58	62	66	70
13	13	17	21	25	29	34	38	42	46	51	55	60	64	68	73	77
14	14	18	23	27	32	37	41	46	51	56	60	65	70	75	79	84
15	15	20	25	30	35	40	45	50	55	60	65	71	76	81	86	91
16	16	22	27	32	38	43	48	54	60	65	71	76	82	87	93	99
17	18	23	29	35	40	46	52	58	64	70	76	82	88	94	100	106
18	19	25	31	37	43	49	56	62	68	75	81	87	94	100	107	113
19	20	26	33	39	46	53	59	66	73	79	86	93	100	107	114	120
20	21	28	35	42	49	56	63	70	77	84	91	99	106	113	120	128

Appendix 2:

Values of K* for finding approximate 95% confidence intervals for

---

Does spironolactone have a place in treating facial hirsutism in women?

Yes it does: the aldosterone antagonist spironolactone has antiandrogenic actions that account for the side effects of gynaecomastia and loss of libido in men and menstrual irregularity in women. It reduces adrenocortical and gonadal synthesis of testosterone and also binds to testosterone receptors in the skin and other tissues, thus reducing the effectiveness of the androgen.[1] The doses used in successful treatment of hirsutism in women have varied from 50 to 200 mg daily. A reduction in hair growth usually occurs within three months. The unwanted effects are diuresis and menstrual disturbance. Diuresis is usually mild and should not cause trouble. Any menstrual disturbance may manifest itself as polymenorrhoea, menorrhagia, oligo-menorrhoea, or amenorrhoea and has been attributed to the antioestrogen and progestational actions of spironolactone. The drug seems to be equally effective in treating idiopathic hirsutism and hirsutism caused by the polycystic ovary syndrome.[2][3] Some women show no improvement, and because of its antiandrogenic effects and its ability to pass through the placenta there is a risk during pregnancy of feminising a male fetus. Spironolactone is particularly useful in treating hirsute women with hypertension in whom treatment with cyproterone and oestrogen is contraindicated.—T E T WEST, consultant physician, Shrewsbury.

1 Loriaux DL, Menard R, Taylor A, Pita JC, Santen R. Spironolactone and endocrine dysfunction. Ann Intern Med 1976;85:630-6.
2 Cumming DC, Yang JC, Rebar RW, Yen SSC. Treatment of hirsutism with spironolactone. JAMA 1982;247:1295-8.
3 Evans DJ, Burke CW. Spironolactone in the treatment of idiopathic hirsutism and the polycystic ovary syndrome. J R Soc Med 1986;79:451-3.